

**IMPROVING THE PEER INTERACTIONS OF STUDENTS WITH
EMOTIONAL AND BEHAVIORAL DISORDERS THROUGH SELF-EVALUATION
PROCEDURES: A COMPONENT ANALYSIS AND
GROUP APPLICATION**

LEE KERN

UNIVERSITY OF PENNSYLVANIA

DAVID P. WACKER

UNIVERSITY OF IOWA

F. CHARLES MACE

UNIVERSITY OF PENNSYLVANIA

AND

GEORGE D. FALK, GLEN DUNLAP, AND JEFFREY D. KROMREY

UNIVERSITY OF SOUTH FLORIDA

We conducted two experiments examining the effects of a self-evaluation package on the peer interactions of students described as emotionally or behaviorally disordered. Experiment 1 assessed the additive effects of various components of a self-evaluation package on the frequency of inappropriate and appropriate peer interactions. The components assessed were rewards alone, rewards plus discussion, and self-evaluation plus rewards. Results showed limited effectiveness when rewards alone and rewards plus discussion were implemented. However, substantial improvements in peer interactions were observed when the self-evaluation component was added. Experiment 2 examined the efficacy and feasibility of the procedures when implemented in a group setting. Students in three classrooms served as participants. Direct observation data collected for 8 of the participants showed the procedures to be effective in improving peer interactions when implemented in a group context.

DESCRIPTORS: self-evaluation, severe emotional disturbance, videotape feedback, social skills

Children described as emotionally or behaviorally disordered frequently have difficulties interacting with their peers in an acceptable manner (e.g.,

Epstein, Kauffman, & Cullinan, 1985; Friedman et al., 1988; Gresham, 1982). One intervention that has resulted in improvements in children's social skills combines self-evaluation procedures with the use of videotape (e.g., Booth & Fairbank, 1984; DeRoo & Haralson, 1971). In a study by Kern-Dunlap et al. (1992), children's peer interactions were improved using a videotape feedback package. The package consisted of self-observation, self-evaluation, and delayed reinforcement. The intervention consisted of daily video feedback sessions during which the children watched videotapes of themselves engaged in a group activity and evaluated the appropriateness of their behavior. The students were rewarded for exhibiting appropriate peer interactions and for correctly self-evaluating their behavior.

Support for this research was provided by The Donald D. Hammill Foundation, Children's Seashore House, and Field-Initiated Research Grant H023C10102 (Office of Special Education Programs). Opinions expressed in this manuscript are those of the authors, and no endorsement by supporting agencies should be inferred.

The authors are grateful for the cooperation and assistance provided by Russell Mitchell, Gilberto Pena, Ken Jones, Denise Hailey, and Sean Casey. The authors also extend appreciation to Jim Paul, Ann Cranston-Gingras, and Bob Friedman for their comments on the design and interpretation of the results.

Reprints may be obtained from Lee Kern at Children's Seashore House, 3405 Civic Center Boulevard, Philadelphia, Pennsylvania 19104.

Although this study was unique in implementing self-management procedures via videotape to improve the peer interactions of students with emotional and behavioral disorders, it is not clear whether the entire package, including self-evaluation, discussion (e.g., feedback, student-generated examples), and rewards, was necessary to obtain the observed behavioral gains. In addition, the intervention was implemented on a one-to-one basis outside of the classroom setting. Such one-to-one interventions are typically not feasible in educational settings. However, the efficacy of the procedures when implemented in a group context remains untested.

The purpose of Experiment 1 was to evaluate whether a simpler package would result in improvements in peer interactions. We assessed the appropriateness of children's peer interactions when they were exposed to rewards alone, to rewards plus discussion, and to a self-evaluation package that consisted of rewards, discussion, and self-evaluation via videotape. The purpose of Experiment 2 was to assess the efficacy and feasibility of implementing the procedures in a group context. We implemented the self-evaluation procedures in three classrooms and evaluated the effects of the package on the students' peer interactions.

EXPERIMENT 1: COMPONENT ANALYSIS

METHOD

Participants and Setting

The participants were 3 boys enrolled in a public elementary school program serving students with emotional or behavioral disorders. The program was located on a regular elementary school campus. The 3 participants were nominated to receive the intervention by their teacher because of specific difficulties with peer interactions.

Brian was 11 years old and was enrolled in fifth grade. According to the Wechsler Intelligence Scale for Children-Revised (WISC-R), he had a full-scale IQ of 84. He was described as severely emotionally disturbed and was placed in special edu-

cation for inattentiveness, impulsivity, poor anger control, and excessive resistance.

Chris was 10 years old and was in the fourth grade. He had a full-scale IQ of 104, measured by the WISC-R. He was labeled emotionally handicapped. His referral behaviors consisted of impulsivity, poor anger control, and excessive resistance.

Andy was 12 years old and attended the fifth grade. His WISC-R IQ score was 100. Aggressive, disruptive, and volatile behaviors resulted in a label of emotionally handicapped and special class placement.

Throughout all phases of the study, daily activity sessions were held in the children's classroom. Eight of 9 students enrolled in the class participated in the activity sessions (consent to videotape was not provided for 1 student). Intervention sessions took place during the rewards-plus-discussion and self-evaluation-plus-rewards phases. These sessions were conducted in a small room adjacent to the school library that was equipped with a videocassette recorder and monitor. Intervention sessions were conducted by one of two facilitators. The facilitators were both graduate students with extensive experience with students with emotional or behavioral disorders.

Behavioral Definitions

Inappropriate and appropriate peer interactions served as the dependent measures. All verbal and nonverbal behaviors directed toward peers were recorded as a peer interaction. A new interaction was scored when (a) 3 s or more elapsed between a child's verbal or nonverbal behaviors, (b) a verbal or nonverbal interchange occurred between the target child and his peer, or (c) verbal or nonverbal behavior was redirected from 1 child to a different child. All peer interactions were classified as either inappropriate or appropriate.

An *inappropriate peer interaction* was scored if any verbal or nonverbal interaction occurred that would be unacceptable in a school or community environment. Inappropriate verbal interactions included statements that were derogatory (e.g., "You're stupid," "You don't know how to play

this game'), statements that ordered a peer to do something in a demeaning or demanding manner (e.g., "Pay attention," "Gimme that," "Put that down"), noises or statements that were spoken in a volume that was noticeably louder than normal and/or that interrupted the ongoing flow of the activity (e.g., screaming, loud singing), and statements that were inappropriate to the context of a school or community environment (e.g., sexual in nature). Inappropriate nonverbal interactions included insulting or derogatory gestures (e.g., sticking one's tongue out).

An *appropriate peer interaction* was defined as any verbal or nonverbal interaction that was acceptable in the context of a school or community environment. Appropriate verbal interactions included statements that were validating, such as praise statements (e.g., "Good job," "You're going to win," "Good try"), statements that were intended to help a peer successfully complete his or her turn in the game being played (e.g., "You get another turn," "Move three more spaces"), and statements or questions that were neutral in nature (e.g., "How many cards do you have left?," "I want to buy Boardwalk," "I like this game, do you?"). Appropriate nonverbal interactions included validating gestures (e.g., a thumbs-up signal) or supportive touches (e.g., high fives, patting a peer on the back, etc.).

Design, Measurement, and Interobserver Agreement

A multiple baseline across students design was used to assess the frequency of inappropriate and appropriate peer interactions while components of the intervention package were added one at a time. The order in which the components were added was based on ease of implementation judged by the classroom teacher. The components evaluated were rewards alone, discussion plus rewards, and self-evaluation plus rewards. All data were collected from videotapes by two observers experienced with both data collection and the behaviors of individuals with emotional and behavioral challenges. The frequency of inappropriate and appropriate inter-

actions was assessed during consecutive 1-min intervals.

Interobserver agreement was assessed during 27% of the sessions distributed evenly across phases and students. Agreement was calculated by dividing the number of agreements per interval by the number of agreements plus disagreements per interval and multiplying by 100%. Mean interobserver agreement for inappropriate peer interactions was 92% (range, 80% to 100%), 92% (range, 69% to 100%), and 91% (range, 83% to 96%) for Brian, Chris, and Andy, respectively. Mean interobserver agreement for appropriate interactions was 90% (range, 83% to 97%), 91% (range, 75% to 97%), and 89% (range, 83% to 96%) for Brian, Chris, and Andy, respectively.

Procedure

Activity sessions were held daily during all phases of the study. During these sessions, the children were given a group game to play (e.g., Uno®, Jenga®, Thin Ice®, Aggravation®, etc.). Activity sessions lasted 12 to 15 min, ending either at the natural conclusion of a game or after 15 min had elapsed.

To simulate a natural or unsupervised setting free from immediate external contingencies, the children were left alone throughout the activity session. The sessions were supervised by one or two adults from a distance of at least 3 m. Neither adult interacted with the children except to clarify the rules of a game or, on two occasions, to intervene when a student became physically aggressive. All activity sessions were videotaped using a portable camcorder stationed on a tripod in the corner of the room. In order to desensitize students to the camera's presence, activity sessions with videotaping were begun 2 weeks prior to initiating baseline.

Baseline. During baseline, activity sessions were carried out in the manner described above. No feedback or contingencies were provided while this condition was in effect.

Rewards. During the rewards phase, each student was instructed that he could earn and select a reward (e.g., pencil, balloon, army figure, etc.)

for reductions in inappropriate behaviors during the activity sessions. Prior to initiating this phase, one of the facilitators held a brief meeting with each student. The students and the facilitator provided examples of inappropriate and appropriate peer interactions. All of the students were able to generate accurate examples and to classify the facilitator's examples correctly. During this phase, rewards were provided contingent on achievement of an individualized reinforcement criterion. The criterion was that inappropriate peer interactions be reduced by at least 70% from the student's mean rate during baseline. This criterion was based on the classroom teacher's judgment of a reasonable behavioral improvement resulting in tolerable levels of inappropriate classroom interactions. Rewards were provided immediately prior to the activity session to students who had earned them during the previous day's session.

Discussion plus rewards. During this phase, the students earned rewards in the same manner as they did during the rewards phase. In addition, each student participated in a 15-min discussion session with one of the two facilitators on the school day following each activity session. During these discussion sessions, the students were asked to describe inappropriate interactions in which they had engaged during the previous activity session. They were also asked to provide examples of more appropriate interactions. The facilitator also described inappropriate peer interactions that occurred during the previous activity session and suggested or asked the students to suggest alternative appropriate interactions.

Self-evaluation plus rewards. During this phase, daily 15-min sessions were again held on the school day following each activity session. During these sessions, the student was shown a total of 5 min of videotape each day divided into 30-s segments. The student was instructed to watch only himself and, when the videotape was stopped, self-evaluate his behavior during the 30-s interval he had just watched by responding to the statement, "I got along with my classmates," with a "yes" or "no" response and recording that response on a

self-evaluation recording sheet. The facilitator simultaneously watched the videotape and also evaluated the appropriateness of the student's behavior. If the student was accurate in evaluating his behavior (i.e., matched the facilitator's response by corresponding with his or her "yes" or "no" rating), the student was awarded 2 points. Further, for each "yes" response that corresponded with the facilitator's "yes," the student earned an additional 3 points. Discussion occurred in the same manner as in the previous phase. For example, when a student's response was "no," he was asked to describe the behavior resulting in that response and to identify what would have been a more appropriate response in that situation. To make the amount of reinforcement the children could earn in this phase comparable to the amount they were able to earn during the rewards and rewards-plus-discussion phases, the items were identical to the previous phase, but were priced so that at least 70% of the points possible during the session needed to be earned to purchase an item.

RESULTS

The results are reported in Figure 1. During the baseline phase, the frequency of inappropriate peer interactions was variable yet high for all of the students. The frequency of appropriate interactions showed a decreasing trend for all of the students. When the rewards phase was implemented, reductions in the frequency of inappropriate interactions were observed initially for Andy; however, an increasing trend was observed while the phase was in effect. Although the variability in Brian's and Chris's inappropriate interactions decreased, the overall frequency did not differ significantly from baseline. None of the students met the 70% criterion during this phase. The frequency of appropriate interactions remained similar to baseline.

When the rewards-plus-discussion phase was implemented, frequencies of inappropriate interactions were observed at levels similar to the baseline phase for all of the students. During this phase, Andy was the only student to earn a reward, on Day 32, by meeting the 70% reduction criterion.

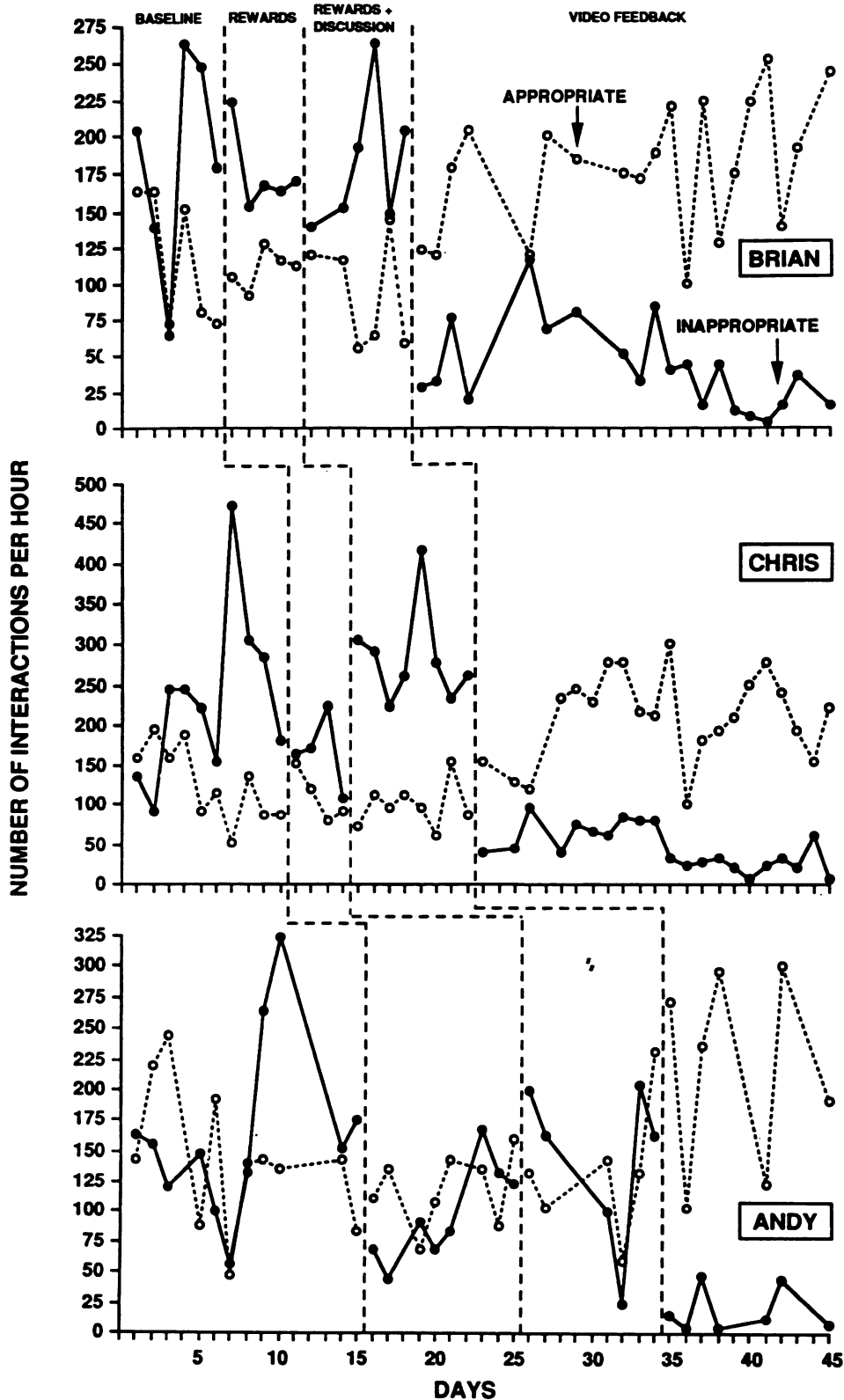


Figure 1. Number per hour of inappropriate and appropriate interactions for each of the students in Experiment 1.

No significant changes were observed in the frequency of appropriate interactions.

Implementation of the video feedback phase resulted in immediate reductions in the frequency of inappropriate interactions. These frequencies remained low while the procedures were in effect. Increasing trends in the frequency of appropriate interactions were observed for all of the students. During this phase, rewards were earned by meeting criterion 67%, 75%, and 100% of the time for Brian, Chris, and Andy, respectively.

Data from the self-evaluation recording forms completed during the video feedback sessions revealed high agreement between the students and the instructor on the occurrence or nonoccurrence of appropriate peer interactions. The mean percentage of agreement for Brian, Chris, and Andy was 87% (range, 56% to 100%), 90% (range, 67% to 100%), and 89% (range, 60% to 100%), respectively.

The results from Experiment 1 show that peer interactions improved only after the self-evaluation component was added to the rewards and discussion components. Experiment 2 was conducted to determine whether this package would be effective and feasible when implemented in a group context. In addition, a sequential withdrawal design was used to determine whether systematic withdrawal of components of the package would result in maintenance of the observed behavioral gains.

EXPERIMENT 2: GROUP APPLICATION

METHOD

Participants and Setting

The participants in this experiment consisted of students enrolled in a partial hospitalization program in eastern Pennsylvania. Each student displayed persistent behavioral challenges both in and outside of school. In addition, all of the students exhibited emotional or behavioral challenges warranting a psychiatric diagnosis and were recipients of medical assistance.

The participating students were enrolled in three

classrooms. Classroom 1 consisted of 3 girls and 6 boys between the ages of 7 and 11 years. Classroom 2 consisted of 10 boys, aged 9 to 13 years. Classroom 3 consisted of 8 girls between the ages of 11 and 13 years. Although all of the students in the three classrooms were exposed to the intervention, for practical reasons data were collected on only 2 or 3 students from each classroom. To select students for data collection, the frequency of each student's inappropriate peer interactions during activity sessions was assessed. The names of students who engaged in moderate to high frequencies of inappropriate peer interactions were pooled. From this pool, the names of 2 or 3 students from each class were randomly selected for data collection.

The students targeted for data collection ranged in age from 8 to 13 years. Results from a variety of intelligence tests revealed a span of intellectual functioning from 75 to 115. The students had received a variety of psychiatric diagnoses and exhibited a range of challenging behaviors. Specific information is reported in Table 1.

The daily activity sessions and the intervention sessions were held in a large room located on the campus. Video equipment on a movable cart was brought into the room during the sessions requiring self-evaluation. The intervention sessions were conducted by a facilitator alone in Classroom 2 (the instructor did not participate for reasons unrelated to the study) and by both a facilitator and the classroom instructor in Classrooms 1 and 3. The facilitator was a graduate student experienced with children with emotional or behavioral challenges.

Behavioral Definitions

Inappropriate and appropriate peer interactions served as the dependent variables in this experiment. Both were defined as in Experiment 1.

Design, Measurement, and Interobserver Agreement

A partial sequential withdrawal design (Kazdin, 1982; Rusch & Kazdin, 1981) was used to assess the effects of the intervention package on the frequency of inappropriate and appropriate peer interactions (self-evaluation plus rewards). This de-

Table 1
Participant Characteristics

Name	Age	IQ	School/psychiatric diagnosis	Referral behaviors
Juan	8	composite = 86 (Kauffman Brief Intelligence Test)	attention deficit disorder with hyperactivity	acting out, hyperactivity
Miguel	8	composite = 88 (Kauffman Brief Intelligence Test)	attention deficit disorder with hyperactivity	poor peer interactions, aggression, poor impulse control, hyperactivity
Tommy	12	V = 73, P = 84 (Bender-Gestalt)	adjustment disorder with work or academic inhibition	poor peer interactions, aggression, hyperactivity
Earl	9	115 (WISC-R)	adjustment reaction with disturbance of conduct	poor peer interactions, aggression, disruption
Larry	12	95 (WISC-R)	adjustment disorder with mixed disturbance of emotions and conduct	poor peer interactions, short attention span, tantrums
Susan	13	75 (WISC-R)	oppositional defiant disorder, attention deficit disorder with hyperactivity	poor peer interactions, fighting
Debby	13	95 (WISC-R)	adjustment reaction with mixed emotional features	poor social interactions
Mara	12	composite = 78 (Kauffman Brief Intelligence Test)	adjustment reaction with disturbance of conduct	tantrums at school

sign also was used to assess maintenance of behavioral gains when various components of the intervention package were withdrawn (the rewards and the self-evaluation procedures) and when the remaining component was faded in Classroom 1. Again, data were collected from videotapes by two individuals experienced with both data collection and individuals with challenging behaviors. Data were collected on the frequency of inappropriate and appropriate peer interactions using Zenith Minisport® lap-top computers.

Interobserver agreement was assessed during 27% of the sessions distributed equally across phases and students. The Reliable computer program (Repp, Harman, Felce, VanAcker, & Karsh, 1989) was used to calculate agreement on frequency of inappropriate and appropriate interactions. Frequency agreement was calculated using a window algorithm whereby an agreement occurred when two observers recorded the occurrence of the same coded behavior within ± 5 s of each other. In Classroom 1, mean interobserver agreement for inappropriate

interactions was 83% (range, 60% to 100%) and 82% (range, 50% to 100%) for Juan and Miguel, respectively. Mean agreement for appropriate interactions was 91% (range, 73% to 100%) and 81% (range, 50% to 100%) for Juan and Miguel, respectively. In Classroom 2, mean interobserver agreement for inappropriate interactions for Tommy, Earl, and Larry was 75% (range, 50% to 100%), 80% (range, 60% to 100%), and 73% (range, 59% to 80%), respectively. For appropriate interactions, mean agreement was 90% (range, 83% to 100%), 86% (range, 71% to 94%), and 82% (range, 72% to 89%) for Tommy, Earl, and Larry, respectively. In Classroom 3, mean interobserver agreement for inappropriate interactions was 85% (range, 58% to 100%), 94% (range, 75% to 100%), and 93% (range, 83% to 100%) for Susan, Debby, and Mara, respectively. Mean interobserver agreement for appropriate interactions was 84% (range, 75% to 100%), 81% (range, 67% to 90%), and 84% (range, 67% to 100%) for Susan, Debby, and Mara, respectively. All of the sessions in which

interobserver agreement was below 66% contained fewer than five occurrences of the target behavior.

Consumer satisfaction and social validation. To provide a measure of consumer satisfaction with the intervention, the classroom instructors involved with implementation of the intervention procedures were asked to respond to seven questions on a 5-point Likert-type scale (available from the first author upon request) with 1 indicating *strongly agree* and 5 indicating *strongly disagree*. These questions called for judgment of the ease with which the procedures were learned, the feasibility with which the procedures could be implemented, and the efficacy of the procedures. In addition, the instructors were given the opportunity to provide comments or suggestions relevant to the self-evaluation procedures.

A semantic differential (available from the first author upon request) was used to measure social validity of the intervention effects. This instrument contained five pairs of antonyms describing student behavior (e.g., well behaved/poorly behaved, good sport/poor sport). Respondents were asked to place a check on one of seven lines between the two adjectives indicating how they judged each student's behavior during randomly selected videotape segments. To assess the internal consistency of this instrument, a Cronbach's alpha (Chronbach, 1951) was conducted using pilot data. Six individuals watched a total of 72 3-min segments of videotape. The Cronbach's alpha yielded a coefficient alpha of .96. The instrument was then administered to 8 individuals naive to the purposes of the study. One child from each class, who was among those targeted for data collection, was randomly selected for social validation assessment. The observers watched a 3-min videotape segment, presented in a counterbalanced order, of each of the 3 children from both baseline and the self-evaluation-plus-rewards conditions. To select representative samples of each child's performance for this assessment, sessions were chosen in which the child's mean frequency of inappropriate interactions during the session was nearest to the condition mean for that child. A two-factor (Child \times Condition) repeated measures ANOVA (Kennedy & Bush, 1985) was

used to analyze the data. To correct for positive bias, a Geisser-Greenhouse conservative *F* test was used when a significant *F* test was obtained.

Procedure

Daily activity sessions were conducted in each classroom in the same manner as in Experiment 1.

Baseline. During baseline, data were collected on the frequency of inappropriate and appropriate interactions as in Experiment 1.

Self-evaluation plus rewards. Following stability in baseline, self-evaluation plus rewards was implemented. Prior to initiating this condition, the instructors in Classrooms 1 and 3 were taught the steps of implementing self-management procedures (Dunlap, Dunlap, Koegel, & Koegel, 1991) during a brief 10- to 15-min meeting with the facilitator. During this phase, all of the students in each class simultaneously watched the videotape from the previous activity session along with either the facilitator alone (Classroom 2) or the facilitator and the classroom instructor. As in the component assessment, the videotape was stopped following each 30-s segment. Matching occurred by comparing the response marked by the instructor and/or the facilitator to the response of 1 or 2 randomly selected students. Points were awarded in the same manner as during the component assessment (i.e., 2 points for correct matching and 3 points for a "yes" response); however, they were awarded to the entire class rather than to the individual student.

At the end of each self-evaluation session, the points earned by the students were totaled. The students in Classrooms 2 and 3 elected to accumulate their points and spend them on pizza parties. Because the students in Classroom 1 were younger, they were given a daily reward (e.g., sticker, piece of candy, baseball card, etc.) following each session in which at least 75% of the possible points were earned.

Rewards only. During the rewards-only condition (Classroom 1), the self-evaluation component was withdrawn and only the rewards component continued to be implemented. The students no longer viewed the videotapes. Instead, the facilitator alone randomly selected the name of 1

student and evaluated that student's performance in the same manner as during the self-evaluation-plus-rewards condition. Five points were awarded to the class for each segment in which the facilitator evaluated the student's peer interactions as appropriate. The students were informed each day of the number of points they had earned during the previous activity session.

Self-evaluation only. During this condition, the reward component was withdrawn while the self-evaluation procedures remained intact. The students continued to view videotapes from the activity sessions and evaluate their performance using the self-evaluation and recording procedures; however, no points or tangible rewards were awarded during this phase.

Fading. The procedures were faded in Classroom 1 to assess the durability of the intervention gains in the absence of daily exposure to the self-evaluation procedures. The activity sessions continued to be conducted 4 days per week; however, the self-evaluation component of the intervention package was faded so that the children evaluated their peer interactions during the activity sessions every other day rather than every day.

RESULTS

The results are shown in Figure 2. During baseline, the students showed variable frequencies of both inappropriate and appropriate peer interactions. The frequency of inappropriate interactions was high overall and exceeded the frequency of appropriate interactions for most of the students throughout this phase. Following implementation of self-evaluation plus rewards, decreases in the frequency of inappropriate interactions were observed for all of the students. The frequency of appropriate interactions either remained stable or showed increasing trends.

When the procedures were withdrawn in Classroom 1, an immediate increase in the frequency of inappropriate interactions was observed with frequencies similar to the initial baseline phase. Reimplementation of self-evaluation plus rewards again resulted in reduced frequencies of inappropriate peer interactions.

In Classroom 1, when the rewards component was withdrawn and only the self-evaluation procedures were in effect, the frequency of inappropriate peer interactions remained low, and the frequency of appropriate interactions remained stable. Low frequencies of inappropriate interactions continued to be observed while the procedures were being faded with Juan. Although limited, the data from this condition indicate that when the self-evaluation component of the package was withdrawn in Classes 2 and 3 and only the rewards remained intact, the students' inappropriate interactions remained low. The frequency of appropriate interactions stayed at levels comparable to those in the self-evaluation-plus-rewards phase.

Because all of the students in the classroom participated in the video feedback sessions, the data on accurate matches from the self-recording forms sometimes reflected few opportunities to assess correspondence. Nonetheless, the data reveal high agreement between the students and the instructor. Larry's, Susan's, and Mara's self-evaluations corresponded with the instructors' evaluations on 100% of the observation intervals. Juan's self-evaluation matched the instructors' evaluations on an average of 78% (range, 50% to 100%) of the intervals during self-evaluation plus rewards and an average of 93% (range, 66% to 100%) of the intervals during self-evaluation only. Miguel's mean percentage agreement was 87% (range, 60% to 100%) during self-evaluation plus rewards. Tommy's correct matches averaged 90% (range, 0% to 100%). The mean agreement for Earl was 94% (range, 50% to 100%) of the intervals. Correct matching for Debby occurred on an average of 90% (range, 50% to 100%) of the intervals.

The results of the assessment of consumer satisfaction are reported in Table 2. The classroom instructors' responses indicated overall satisfaction with the procedures. Comments offered by the teachers also suggested satisfaction with the procedures (e.g., "The ability to see yourself and have to say, 'That's me being negative towards my peers,' assisted in their growth").

The results from the social validation assessment of intervention effects showed that observers naive

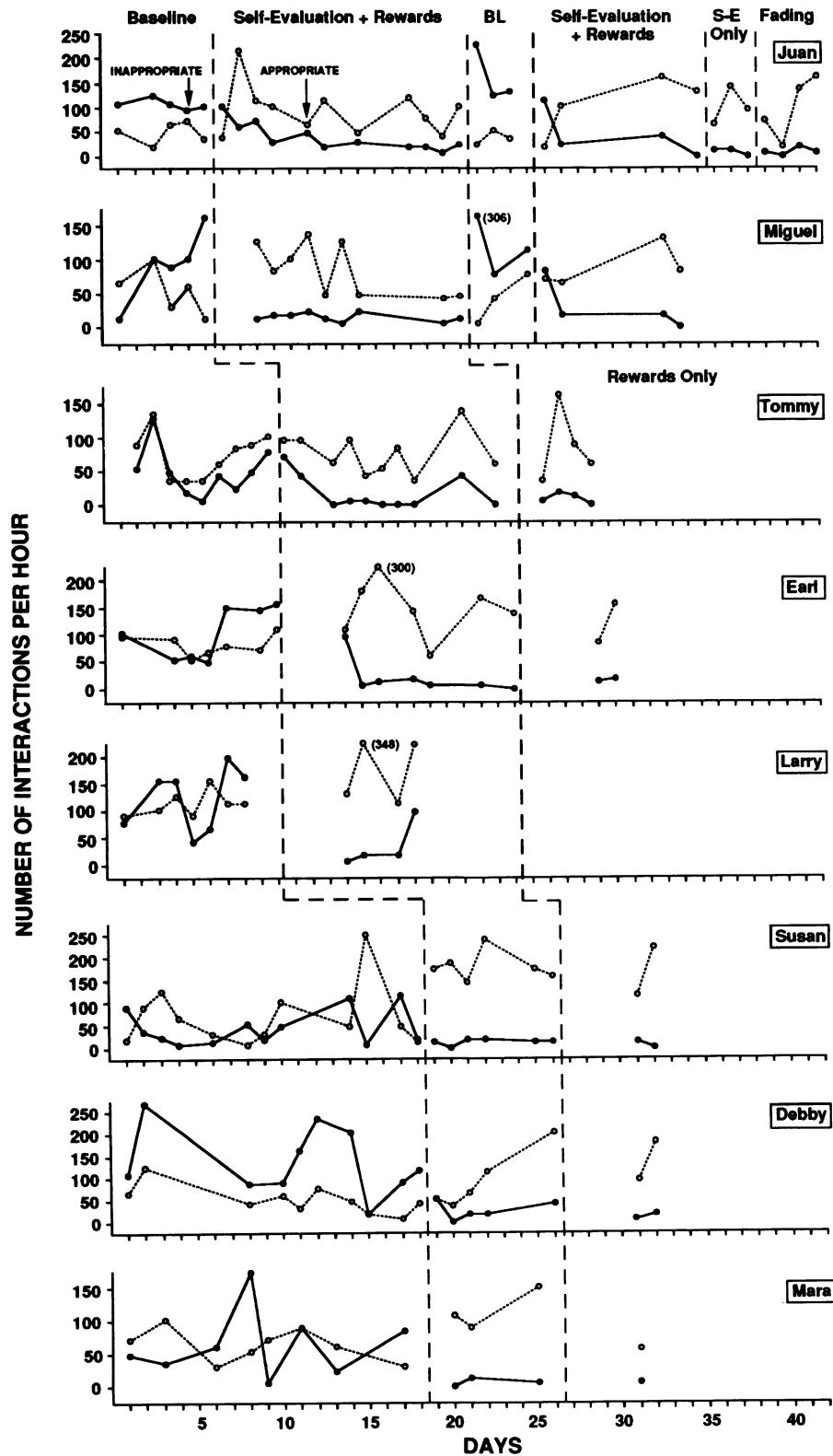


Figure 2. Number per hour of inappropriate and appropriate interactions for each of the students in Experiment 2.

Table 2
Instructors' Responses to Social Validity Questionnaire

Statement	Response	
	Instructor 1	Instructor 2
The self-evaluation procedures were easy to learn.	1	1
The self-evaluation procedures were difficult to implement with my students.	5	5
The procedure took a reasonable amount of time to implement in the classroom.	2	5
The procedures were effective in improving my students' peer interactions during the activity sessions.	1	1
The procedures were effective in improving my students' peer interactions throughout the day.	2	1
Other social skills procedures I have used have produced better results than the self-evaluation procedures.	5	3
I prefer using the self-evaluation procedures than other social skills procedures.	1	3

Note. 1 = strongly agree; 5 = strongly disagree.

to the purposes of the study rated the children's behavior significantly better during the self-evaluation-plus-rewards phase than during baseline. The results of the two-factor repeated measures ANOVA, using the Geisser-Greenhouse conservative *F* test, revealed a main effect for condition, $F(1, 37) = 15.54, p < .006$.

GENERAL DISCUSSION

The results from this study are consistent with previous research showing the effectiveness of self-evaluation using videotape procedures (Kern-Dunlap et al., 1992) for improving peer interactions in children with emotional and behavioral disorders. The findings from the component assessment suggest that improvement in peer-to-peer interactions may be accomplished best with the implementation of a package consisting of both rewards and videotape feedback, rather than rewards or rewards and discussion alone. The results of the group application establish the effectiveness and feasibility of the procedures when implemented in a group context. This study represented a departure from earlier analyses by evaluating the combined contribution of various components of the intervention package and by demonstrating the effectiveness of the procedures when implemented in a group setting.

As suggested by Kern-Dunlap et al. (1992), the effectiveness of the package may be best explained by the concept of rule-governed behavior (e.g.,

Poppen, 1989; Reece, 1989). This speculation is supported, in part, by the rapid behavioral improvements following implementation of the package and the high accuracy of the children's self-evaluations. It is also possible that peer-delivered reinforcement and punishment, observed anecdotally during the activity sessions, contributed to the observed behavioral gains. Finally, it is also likely that the use of videotape played a seminal role in the effectiveness of the procedures. The videotape allowed the procedures to be highly individualized, provided a permanent and accurate record of the children's behavior, and allowed evaluation to occur without conflicts.

Of additional interest was the observed maintenance of behavioral improvements following withdrawal of components of the intervention package in the group application. Although the results from the component analysis suggest that inclusion of the self-evaluation procedures may be necessary for initial acquisition, the results from the group application suggest that the entire package may not be necessary to maintain behavioral gains. It is possible that the partial sequential withdrawal design enhanced maintenance because removal of the contingencies were less discriminable to the participants. This hypothesis is partially supported by the results observed during the complete withdrawal of the package conducted in Classroom 1. When the entire package was withdrawn, high frequencies of inappropriate interactions were observed. This suggests that component withdrawal

was less obtrusive and may have enhanced maintenance. It is also possible that natural contingencies (i.e., punishment and reinforcement) provided by peers, as described earlier, became operative in the course of the study.

A few caveats should be noted. First, an additive model was used in Experiment 1. Thus, the effect of self-evaluation alone was not assessed. It could be argued that this component alone may have resulted in the observed behavioral improvements. However, the results of previous research (Kiburz, Miller, & Morrow, 1984) suggest that self-management procedures in the absence of reinforcement have limited efficacy for initial acquisition.

Another possible confounding effect in Experiment 1 pertains to the administration of rewards across phases. The addition of the self-evaluation procedures during the last phase produced distinctive differences from the previous two phases. Thus, it was not possible to keep the contingencies for behavior change identical. Nonetheless, data indicate that students met or exceeded the 70% behavioral reduction criterion in effect during the previous two phases 88% of the time during the video feedback phase. When replicating this study, it may be necessary to modify the proportion of points awarded for accuracy if accurate self-evaluation is occurring in the absence of behavioral improvements.

Although the procedures used in this study were effective in improving the children's peer interactions during the activity sessions, no data were collected in other settings. Thus, it is not possible to evaluate whether, or to what extent, the procedures resulted in generalized improvements in the children's peer interactions. In spite of the arrangement of the activity sessions to simulate situations typical of those the children might naturally encounter, events such as the presence of the video-camera, may have served as discriminative stimuli signaling a time in which the children's behavior would be monitored.

The context in which this study was conducted may limit the applicability of the findings. Both the component analysis and the group application were conducted in nonintegrated settings. There is

evidence that frequency of inappropriate peer interactions is higher when children interact only with peers who share similar difficulties (e.g., Brinker, 1985; Halvorsen & Sailor, 1990; Won, Anderson, & Haring, 1993). It is possible that different results might accrue if the interventions were implemented in an integrated context.

The results of this study suggest an effective method for improving the social performance of children who are emotionally disturbed. Further research might address issues related to generality of the results. In addition, the findings raise a number of questions involving conceptual interpretations, such as the mechanisms underlying the effectiveness of self-management procedures.

REFERENCES

- Booth, S. R., & Fairbank, D. W. (1984). Videotape feedback as a behavior management technique. *Behavioral Disorders*, 9, 55-59.
- Brinker, R. P. (1985). Interactions between severely mentally retarded students and other students in integrated and segregated public school settings. *American Journal of Mental Deficiency*, 89, 587-594.
- Chronbach, C. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- DeRoo, W. M., & Haralson, H. L. (1971). Increasing workshop production through self-visualization on videotape. *Mental Retardation*, 9, 22-25.
- Dunlap, L. K., Dunlap, G., Koegel, L. K., & Koegel, R. L. (1991). Using self-monitoring to increase independence. *Teaching Exceptional Children*, 23, 17-22.
- Epstein, M. H., Kauffman, J. M., & Cullinan, D. (1985). Patterns of maladjustment among the behaviorally disordered II: Boys aged 6-11, boys aged 12-18, girls aged 6-11, and girls aged 12-18. *Behavioral Disorders*, 10, 125-135.
- Friedman, R. M., Silver, S. E., Duchnowski, A. J., Kutash, K., Eisen, M., Brandenburg, N. A., & Prange, M. (1988). *Characteristics of children with serious emotional disturbances identified by public systems as requiring services*. Tampa, FL: Florida Mental Health Institute, University of South Florida.
- Gresham, F. M. (1982). Misguided mainstreaming: The case for social skills training with handicapped children. *Exceptional Children*, 48, 422-430.
- Halvorsen, A. T., & Sailor, W. (1990). Integration of students with severe and profound disabilities: A review of research. In R. Gaylord-Ross (Ed.), *Issues and research in special education* (Vol. 1, pp. 110-172). New York: Teachers College Press.
- Kazdin, A. E. (1982). *Single-case research designs*. New York: Oxford University Press.

- Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the design and analysis of experiments in behavioral research*. New York: University Press of America.
- Kern-Dunlap, L., Dunlap, G., Clarke, S., Childs, K. E., White, R. L., & Stewart, M. P. (1992). Effects of a videotape feedback package on the peer interactions of children with serious behavioral and emotional challenges. *Journal of Applied Behavior Analysis*, 25, 355-364.
- Kiburz, C. S., Miller, S. R., & Morrow, L. W. (1984). Structured learning using self-monitoring to promote maintenance and generalization of social skills across settings for a behaviorally disordered adolescent. *Behavioral Disorders*, 10, 47-55.
- Poppen, R. L. (1989). Some clinical implications of rule-governed behavior. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 325-357). New York: Plenum.
- Reece, H. W. (1989). Rules and rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 3-84). New York: Plenum.
- Repp, A. C., Harman, M. L., Felce, D., VanAcker, R., & Karsh, K. L. (1989). Conducting behavioral assessments on computer collected data. *Behavioral Assessment*, 2, 249-268.
- Rusch, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis*, 14, 131-140.
- Won, K. P., Anderson, J. L., & Haring, K. (1993). *The effects of integration on the challenging behavior of students with severe disabilities*. Unpublished manuscript.

Received February 2, 1994

Initial editorial decision April 11, 1994

Revisions received July 19, 1994; August 22, 1994

Final acceptance September 2, 1994

Action Editor, Brian K. Martens